

# **Analysis and Prediction of User Behaviour in a Museum Environment**

A thesis presented

by

Karl Grieser

to

The Department of Computer Science and Software Engineering

in partial fulfillment of the requirements

for the degree of

BSc (Hons)

University of Melbourne

Melbourne, Australia

October 2006

Thesis advisor(s)  
**Steven Bird**  
**Timothy Baldwin**

Author  
**Karl Grieser**

## **Analysis and Prediction of User Behaviour in a Museum Environment**

### **Abstract**

Visitors to a museum enter an environment with a wealth of information. However not all of this information may be located in physical form. It may be accessible through an online web-site and available for download, or this information could be presented by a tour guide that leads you through the museum. Neither of these sources of information allow a visitor to deviate from a set path. If a visitor leaves a guided tour, they will not have access to the resource that is supplying them with the extra information. If they deviate from a downloaded tour, they again will not have the correct information sheets for an exhibit that is not directly on their tour. The solution is to create a Recommender System based on the conceptual similarity of the exhibits. This system creates a dynamic tour through the museum for a given visitor by recommending exhibits that the visitor is interested in. Conceptual similarity of exhibits can be comprised of elements including the physical proximity, the semantic content of the exhibit, and the opinions of previous visitors. By using a combination of these similarities, we have produced a system that recommends relevant exhibits in 51% of test cases.

# Contents

Title Page . . . . .	i
Abstract . . . . .	ii
<b>1 Introduction</b>	<b>1</b>
1.1 Task Background . . . . .	1
1.2 Methodology . . . . .	2
1.3 Overview . . . . .	3
<b>2 Background</b>	<b>4</b>
2.1 The Digital-Physical Divide . . . . .	4
2.2 Content Extraction, and Prediction . . . . .	5
2.3 Museum-Visitor Interaction . . . . .	7
<b>3 Resources</b>	<b>9</b>
3.1 Museum Structure . . . . .	9
3.2 Melbourne Museum . . . . .	10
3.2.1 The Australia Gallery . . . . .	11
3.2.2 Collected Visitor Data . . . . .	11
3.2.3 Museum Website . . . . .	13
3.3 Natural Language Resources . . . . .	13
3.3.1 WordNet . . . . .	14
3.3.2 WordNet::Similarity . . . . .	14
3.3.3 Lingpipe . . . . .	15
<b>4 Methodology</b>	<b>16</b>
4.1 Physical Location . . . . .	17
4.2 Collaborative Data . . . . .	19
4.2.1 Naive Bayes Prediction . . . . .	20
4.3 Semantic Content . . . . .	20
4.3.1 Word Sense Disambiguation . . . . .	21
4.3.2 WordNet Similarity and Relatedness . . . . .	21
4.4 Evaluating Component Effectiveness . . . . .	23

<b>5 Experiments and Evaluation</b>	<b>25</b>
5.1 Experiment Design . . . . .	25
5.2 Single Component Predictive Methods . . . . .	26
5.2.1 Sequential Prediction . . . . .	26
5.3 Multiple Component Conceptual Models . . . . .	27
5.3.1 Improving Precision through Thresholds . . . . .	28
<b>6 Conclusion and Future Work</b>	<b>31</b>
6.1 Future work . . . . .	32

# Chapter 1

## Introduction

Portable technology such as mobile phones and Personal Digital Assistants (PDAs) are becoming the norm in everyday life. This technology is slowly but inexorably being adopted by all sectors of society. The major reason for this adoption of technology is its ability to enrich our lives through availability and enhanced communication, and to help co-ordinate our lives through the use of calendars and diaries. However, with few exceptions, these technologies do not take into account how the physical world and information contained in it affect our decisions and actions. Knowledge of how users relate to an information rich environment, and the way they respond to it, can be used to guide and supply information on topics that are relevant to the user's behaviour within a related physical environment. Information about surrounds can be categorized and streamlined to enhance a user's knowledge of a locale, allowing them to make more informed decisions, or to provide the user with detailed additional content that is not normally accessible in a physical space. This research examines the use of heterogeneous content to enhance a user's experience of an environment.

### 1.1 Task Background

The use of non-physical information to provide added meaning to physical environments has been developed as part of large scale research projects aiming to explore the relationships between the physical and digital worlds. Two notable projects are Equator<sup>1</sup> (an Interdisciplinary Research Collaboration (IRC) aiming to bridge the divide between the physical and digital worlds), and the HIPS (Hyper-Interaction within Physical Space) project(Benelli *et al.* 1999). Both of these projects analyse the way in which digital content and virtual representations of physical environments can enhance the way that people live, or perform daily tasks. A key aspect of this integration of technology is the issue of content adaption and presentation. ILEX(Hitzeman *et al.* 1997) is a virtual museum that tailors information to a spe-

---

<sup>1</sup>The Equator Project Website: <http://www.equator.ac.uk/>

cific visitor's preferences. Using this method, visitors are able to access content that they find interesting, and have their own personalised guide tour through the virtual museum.

Guided tours have long been available for visitors to museums, and tourists in cities. The tour guides give museum visitors and tourists additional information about the surrounding environment which they would not normally have knowledge of. Museums such as the Louvre in Paris also give visitors the opportunity to construct their own tours before they visit the museum.<sup>2</sup> The Louvre uses predefined Thematic Trails that take the form of web-pages that can describe the background of exhibits, and define the route taken. The visitors can print off these pages or load them onto a portable device to browse as they tour the museum. The problem with the Louvre's Thematic Trails and a tour guide in a city is that they are statically defined. Once the visitor or tourist has made a choice to follow a Trail or a tour guide they do not have extra content available for other exhibits or locations not directly associated with the tour. If a tourist wishes to explore a location outside the scope of a guided tour, they have to leave the tour group and will not have a guide to tell them about the location they wish to visit. If a museum visitor wants to examine an exhibit outside the scope of their Thematic Trail, they will have to leave their path. They are able to resume their tour at any time, but they will not have the same level of information available as with the exhibits on the pre-made tour.

## 1.2 Methodology

How users relate actions and previously visited locations to one another indicates how they conceptualise connections and relationships between the locations. Patterns that people follow can help to indicate what future actions they might take. These future predictions can be used in all manner of applications, from personal scheduling to presentation and dissemination of information content. For this purpose, I have chosen to work with the Melbourne Museum to create a system to recommend related exhibits to visitors to the museum. In this case the visitor would be presented with a list of exhibits that they may like to visit, based on the previous exhibits they have seen. Other relevant information about the exhibits can also be presented to the visitor to enhance their visiting experience. This research aims to form dynamic recommendations to a museum visitor in order to create a personalised tour based on visitor interests. This will eliminate the need for static pre-made paths and allow the visitor to create their own tour as they move around the museum.

An environment such as a museum offers an advantage over other information rich environments such as a city because of size: a smaller environment is easier to deal with, but also because of the nature of the data within the museum. The

---

<sup>2</sup>Paths of Discovery: [http://www.louvre.fr/llv/activite/liste\\_parcours.jsp?bmLocale=en](http://www.louvre.fr/llv/activite/liste_parcours.jsp?bmLocale=en)

heterogeneity of a curated information rich environment allows effective analysis of how users interact with it, without being hindered by patchy data, or missing content. A museum can also be divided into sub-sections. These include exhibitions, exhibition areas, and individual exhibits. A hierarchical structure such as a museum and its exhibitions is referred to as a taxonomy or ontology.

This research explores the notion of conceptual relationships between exhibits through the similarity of semantic content, and offers a comparison of different measures of relatedness and similarity over WordNet (Fellbaum 1998). Also studied are the effectiveness of simple physical attributes of the exhibits in determining conceptual relatedness. The effect of group training methods versus individual histories as a source of prior path data as an effective means of prediction is also analysed. Specifically, this project aims to determine whether users make use of a conceptual, semantically based, model when touring an information rich environment, and whether accurate predictions can be made based on prior paths. This research shows that users use semantic content associated with exhibits they encounter in order to make decisions on what to visit next.

### 1.3 Overview

Multiple aspects of computing, user modelling, and even architecture, make up the background of this research. We will begin with an in depth discussion of how these fields of research relate to one another is a necessary precursor to the creation of any recommendation system (Chapter 2). This discussion will lead into a detailed analysis of the data to be used in this research, as well as the environment it describes. The environment must be considered carefully in the way that it relates to the data that describes it (Chapter 3). Visitors' conceptual models are comprised of multiple aspects of the surrounding environment. The development of these attributes of both physical and semantic spaces into a set of conceptual models designed to create pertinent recommendations is the core contribution of this research (Chapter 4). The outcomes of experimentation are evaluated to determine the validity of these conceptual models (Chapter 5), and the contribution of this research to future work in the field is analysed (Section 6.1).

# Chapter 2

## Background

There currently exist multiple projects around the world that are investigating the integration between physical and digital worlds. These projects address many issues that arise from such integration: environmental awareness, social interaction, content adaption, and personalisation. Examples of content adaption and personalisation include tailoring presented information to a user based on their age or knowledge background, and remembering user preferences to create recommendations restaurants the user might like. The interaction a user has with their surrounding environment (either physical, or digital) can supply important information about how the user acts in certain situations. This can be as simple as saying a user turns right at intersections 82% of the time, hence recommend that they always turn right. Or more complex recommendations such as anticipating when to buy stock in a particular company.

### 2.1 The Digital-Physical Divide

The Equator Project is an IRC formed by eight academic institutions in the United Kingdom. In broad terms the goal of Equator is to address and evaluate all challenges that the integration of digital and physical worlds poses. The digital world is the unseen space encompassed by digital technology. This includes resources like the Internet, as well as digital technology that is integrated into objects within the surrounding environment (such as digital billboards, interactive department store directories and, more broadly, Global Positioning Systems). Equator was conceived to address the social issues associated with large scale integration of digital technology, as well as the technical challenges. Its key contribution to the field has been to examine how the physical and digital worlds interact with one another. It is this aspect that our research examines: analysis of how an information rich non-physical space can be used to impart meaning onto objects within a physical space.

The term hypernavigation describes the navigation of a virtual environment. Such a representation can allow a virtual representation of a physical space, and hence the

navigation of a physical space by proxy. The HIPS project (Benelli *et al.* 1999) extends this concept one step further, into the realm of hyper-interaction. Hyper-interaction describes the overlap between the physical environment, and the digital information content describing it. Certain actions in the real world will translate into actions in the digital world. For example moving closer to a real world object will display information associated with the that object. The overall aim of the HIPS project was to create a digital tour guide that would be able to recommend objects within the local physical environment and provide information about them. This research aims to analyse the importance that different components of objects have on the effect of recommendations. Specifically this research identifies the key aspects of real world object content to examine when creating recommendations within an information rich space.

A key component in predicting a visitors future movement is the ability to keep track of their current movement. This history must be up to date for any predictions based on it to be valid. If a visitor has just seen something that we are about to recommend, then recommending it is useless. Thus the need for a system that can quickly and accurately track visitor movement is an essential component in creating a prediction system. The most ideal method of tracking visitors is to give each visitor a wireless device that can have its position relayed back to a central server (as in Benelli *et al.* (1999)). Sparacino *et al.* (1999) suggests several methods of using wireless devices to enhance the visiting experience in addition to tracking the visitors through the museum. The two major methods suggested are fixed “Smart Rooms,” which uses a restricted physical space to present non-physical content or to allow visitor interaction, and wearable computers which act as a means for extra audio as well as video content to be presented to the user. These methods Such an implementation is out of the scope of this research. This study deals with path prediction, not locational awareness and visitor tracking. We assume that all tracking is done implicitly, and that visitor paths are known in advance. However, it is necessary to acknowledge the different ways in which visitor are able to interact with a physical environment while relating back to an associated virtual space.

## 2.2 Content Extraction, and Prediction

Visitors to an Information Rich space have an abundance of information to deal with, whether it be advertising, or other content they wish to access. The notion of information being “pushed” onto visitors, and being “pulled” by visitors was explored in Cheverst and Smith (2001). With respect to a museum environment there is an ever present level of information push from all exhibits within the museum. But individual sections were pulled by visitors that approached the exhibits and examined their content. This is one of the aspects that this research will address. By predicting and recommending exhibits to visitors, less time will be lost by visitors trying to work out

which exhibits are pushing desired content, and can go directly to the exhibits they wish to pull content from. Ascertaining what information a visitor is trying to pull from an exhibit can be modelled as determining relationships between exhibits. If a visitor examines two exhibits (i.e. pulls information from both of them), we can use the information overlap to determine what information the visitor pulled from the exhibit. This overlap can then be used to find exhibits with a similar information content, and selective supply the information so that only the desired content is pushed. While this research only deals with recommendation based on exhibit similarity, future studies could further examine this method of content adaption.

Content adaption was also explored notably as part of the Intelligent Labelling Explorer (ILEX) project (Hitzeman *et al.* 1997). In this case the content adaption was more focussed on Natural Language Generation, based on visitor exhibit preferences. A virtual museum in the form of a website was developed to allow users to navigate to any areas or exhibits that they expressed interest in. These navigational histories were used to identify aspects of the exhibits that the users found interesting. These aspects were then related to other exhibits and referenced against a database of facts about each exhibit. Text was then generated to explain connections, unseen aspects, or more detailed content(if desired). However this positive content adaption did not allow users to remove content if they inadvertently strayed towards an uninteresting section. The discourse produced for each exhibit is logged so as not to repeat information, so in this respect all information about an exhibit will be new for each time the visitor sees the exhibit, and even if the user disliked the exhibit it will be related to aspects of other exhibits that they may have enjoyed. This again falls under the category of content adaption, rather than content recommendation, but the methods described here dealing with uncovering unexpected relationships can be extrapolated to a recommendation system. automatically scoring overlap between all exhibits, even seeming unrelated exhibits can be recommended to a visitor (who will then have a nice surprise of uncovering an hidden fact).

Recommender Systems (Resnick and Varian 1997) have the purpose of using a statistical or histographic model to predict user paths. Some well known recommender systems include websites such as del.icio.us,<sup>1</sup> and the internet radio stations Last.fm<sup>2</sup> and Pandora.com.<sup>3</sup> All of these systems attribute a set of identifying keywords or phrases to the objects being recommended (web-pages, songs). These terms are then used to gauge similarity between songs or web pages. Whereas ILEX does not take into account the order of such a history, other studies have analysed the effect of ordered history on Recommender Systems. Such a system was proposed in Chalmers *et al.* (1998) at the seventh World Wide Web Conference as a solution to the relative dearth of effective search engines at the time.<sup>4</sup> The order of a users history is impor-

---

<sup>1</sup><http://del.icio.us>

<sup>2</sup><http://www.last.fm>

<sup>3</sup><http://www.pandora.com>

<sup>4</sup>Interestingly enough, this paper was presented at the same conference as Sergey Brin and

tant, as visitors are less likely to remember exhibits long ago in their path, and are more likely to make decisions based on their current context. This research will examine methods of sequential prediction when making recommendations. The length of the path will also be taken into account when performing sequential predictions.

A user without a history can be managed in several ways: we can wait until they do something, and base predictions on that, or we can assume that the visitor will act in the same way as previous visitors have, and hence use previous visitor paths to create a model. The collaborative approach relies on there being a sufficient amount of previous visitor data to create an (at least partially) accurate classifier for new visitors to the museum. The individual user approach on the other hand can be much better at correctly predicting future movements, but this accuracy is based on the size of the history, the more training data the better. The two methods also allow for different methods of prediction to be used. The collaborative methods knows the most likely future patterns based on examples it has seen. The individual method on the other hand must make judgements based on current data, about future elements it has not seen before. Zukerman and Albrecht (2001) gives a good discussion of the relative drawbacks and advantages of these two approaches. As well as predictive methods and statistical models available in these two classes for the purpose of user modelling. The paths of prior visitors can be used as a means of supplying collaborative recommendations to new visitors. Collaborative as well as individual methods of exhibit recommendation will be used in this research.

## 2.3 Museum-Visitor Interaction

As with any physical space, the space itself provides a restriction on the visitor within that space. In the case of museums, this restriction can be because of the layout of the exhibits themselves, such that the visitor must follow a set path (Bitgood and Shettel 1997). In this case any predictions based on a collaborative method would be useless, as they would all follow the same path (accurate, but useless), and any predictions based on individual user history would just cause the visitor to go backwards and forwards along the same route, thus passing most exhibits multiple times. Studies such as Peponis *et al.* (2004) have shown that for museums that have a sufficiently open plan environment, such effects are minimised. However Peponis *et al.* acknowledge that analysis of visitor paths is needed to identify how any features of a museum exhibition effect visitor motion.

Personalisation: does more feedback help or hinder. By tailoring a tour to a visitor, we are personalising their visiting experience. Such personalisation can be done beforehand by the visitor (as in Filippini (2003), or the Louvre's Thematic Trails), or the content can be dynamically created to a visitor's preference (Hitzeman *et al.* 1997). However this is done, the behaviour of the user changes if they have the

---

Lawrence Page's definitive paper describing Google (Brin and Page 1998)

knowledge that they are under observation, as mentioned in (Chalmers *et al.* 1998; Chalmers 2001). When any recommendation is made to a museum visitor, they will be temporarily aware that they are being tracked. This intrusion is minimal and temporary and, if done well, will go unnoticed. In the case where feedback is necessary for a predictive model to function, this feedback should be as minimal as possible.

# Chapter 3

## Resources

Delivering non-physical information to people that are within a physical space requires a rich source of information to draw from. It also requires a physical space that corresponds with this information. In an environment such as a city this information can be patchy. Not all objects and locations have a corresponding body of information associated with them. It may be easy to find a description of the significance of an office building (date built, services provided, etc.), however finding information relating to a less well known object (an obscure statue in a park for example) can be more difficult.

This chapter describes the sources of the data to be used in this research, as well as the methods used to analyse this data. The environment in which the testing takes place is important to consider carefully. The environment must be a visitable space: people must want to visit it, or else recommending objects within it would be useless. The environment must also be associated with a rich information source that describes it. This information source must be readily accessible, and be of a consistent quality.

This data must be processed and analysed. We are aiming to deliver content to visitors, and using information rich content to enhance the visitor's museum experiences. In order to extract semantic content from bodies of text, the semantic content needs to be identified. This content can be extracted using language processing methods such as Named Entity Detection. A versatile language technology toolkit will also enable testing and evaluation over the collected data.

### 3.1 Museum Structure

The homogeneity of the information within a given space is an import factor when considering an appropriate environment. A consistent standard of information content across an environment ensures an even distribution of the environment's semantic representation. This simply means that the information is not clumpy or patchy

when placed into a physical environment.

Museums are physical locations that provide homogeneous, information rich environments. The information contained within museums is of a consistent standard of quality, and is constantly kept up to date. The consistency of the information style is necessary to avoid any patchiness in the overall content.

Museums are traditionally classified into levels of granularity: exhibitions, exhibition areas, and exhibits. Exhibitions cover large sections of museums and group together all exhibits revolving around a broad theme. Exhibition areas are smaller collections of exhibits, and group exhibits based on a more specific or concise theme. The lowest level of granularity in a museum is an exhibit. An exhibit may be one individual item, such as a painting or a dinosaur bone, or a collection of objects such as several pots or hats. This research deals with exhibits as the objects being recommended, but the methods shown in this research can be used to recommend higher levels of representation (such as exhibition areas) in the museum.

## 3.2 Melbourne Museum

Delivering non-physical adapted content to a user within a physical space necessitates the acquisition of an information rich real world environment, i.e an environment in which real world entities have information describing them that can be easily acquired and accessed. This information can take the form of multimedia such as audio or video, images, or simply descriptive text that describes physical entities is easily accessible. This project was initially proposed as part of a collaborative project with the Melbourne Museum, and as such we have access to the facilities at the Melbourne Museum, as well as data collected by the marketing department within the museum.

The Melbourne Museum is a recently constructed museum located in Carlton Gardens. The museum houses a broad showcase of exhibits compiled into exhibitions on the Natural world, Science and Technology and Australian History. These individual exhibitions are also sub-divided into smaller exhibition areas. The museum is recently built, and as such does not suffer from physical restrictions that can be created by using older buildings not suited to the task of housing a museum. In the case of the Melbourne Museum's previous residence at the State Library of Victoria,<sup>1</sup> visitors were funnelled through corridors, and passed exhibits in a designated order. A rigid museum design does not allow visitors the ability to selectively access pertinent exhibits. The open plan design at Melbourne Museum's current residence allows visitors to have knowledge of the existence of all exhibits within the current exhibition or gallery. As a consequence they are able to make decisions about which exhibits to visit without being restricted by physical constraints aside from distance.

---

<sup>1</sup><http://www.slv.vic.gov.au>

### 3.2.1 The Australia Gallery

The Australia Gallery showcases a range of exhibits that revolve around the central themes of Melbourne's culture and history. It contains exhibits such as CSIRAC,<sup>2</sup> and the taxidermal remains of Phar Lap (Australia's most famous race horse).

The exhibitions within this gallery are varied in topic and content, and offer a sample distribution of different content types offered by a standard museum (exhibits for all ages, knowledge bases and backgrounds). The exhibits are the lowest level of granularity in a museum environment. The exhibits in the Australia Gallery are organised into exhibit areas. These exhibit areas are organised into themes revolving around a specific topic or common element. These themes include the life of Phar Lap, the founding of Melbourne, and Sport in Melbourne. The grouping of similar exhibits into close physical proximity contributes to the overall similarity displayed between exhibits within the area.

For the purposes of experimentation, we define the exhibits as the locations represented in Figure 3.1. These locations each correspond to an exhibit. There are fifty three exhibits in total.

### 3.2.2 Collected Visitor Data

In order to analyse the movement of visitors through galleries, and the museum as a whole, the Marketing and Research Department at the museum routinely follows visitors through the museum and makes detailed recordings of their movements. Statistics such as the path the visitor takes, the time spent at each exhibit, and how they entered and exited the gallery/exhibition are recorded. The method used to track the visitor paths is simple, but accurate. Museum staff use tracing paper layered over a map of the gallery or exhibition, and hand draw in the path that the visitor takes, noting the direction taken and the locations of all stops made. The time that a visitor spends at each individual exhibit is recorded separately and placed into a Microsoft Excel spreadsheet. The traced visitor paths are primarily used to study how the layout and exhibit placement in the museum affects the flow of human traffic. The recorded statistics are used to analyse which exhibits are being visited more or less frequently than others, and why.

The data was collected during September and October of 2001. The age of the data collection must be taken into account, as some exhibits have been added or removed during this time. However this is only the case for two of the exhibits within the Australia Gallery. One exhibit has been added and one removed. The two exhibits are in a very similar position though, and for all intents and purposes can be considered to be the same exhibit. The content of other exhibits has not changed significantly over time, and the semantic content associated with the museum space

---

<sup>2</sup>Council for Scientific and Industrial Research Automatic Computer: The world's fourth computer, and Australia's first.

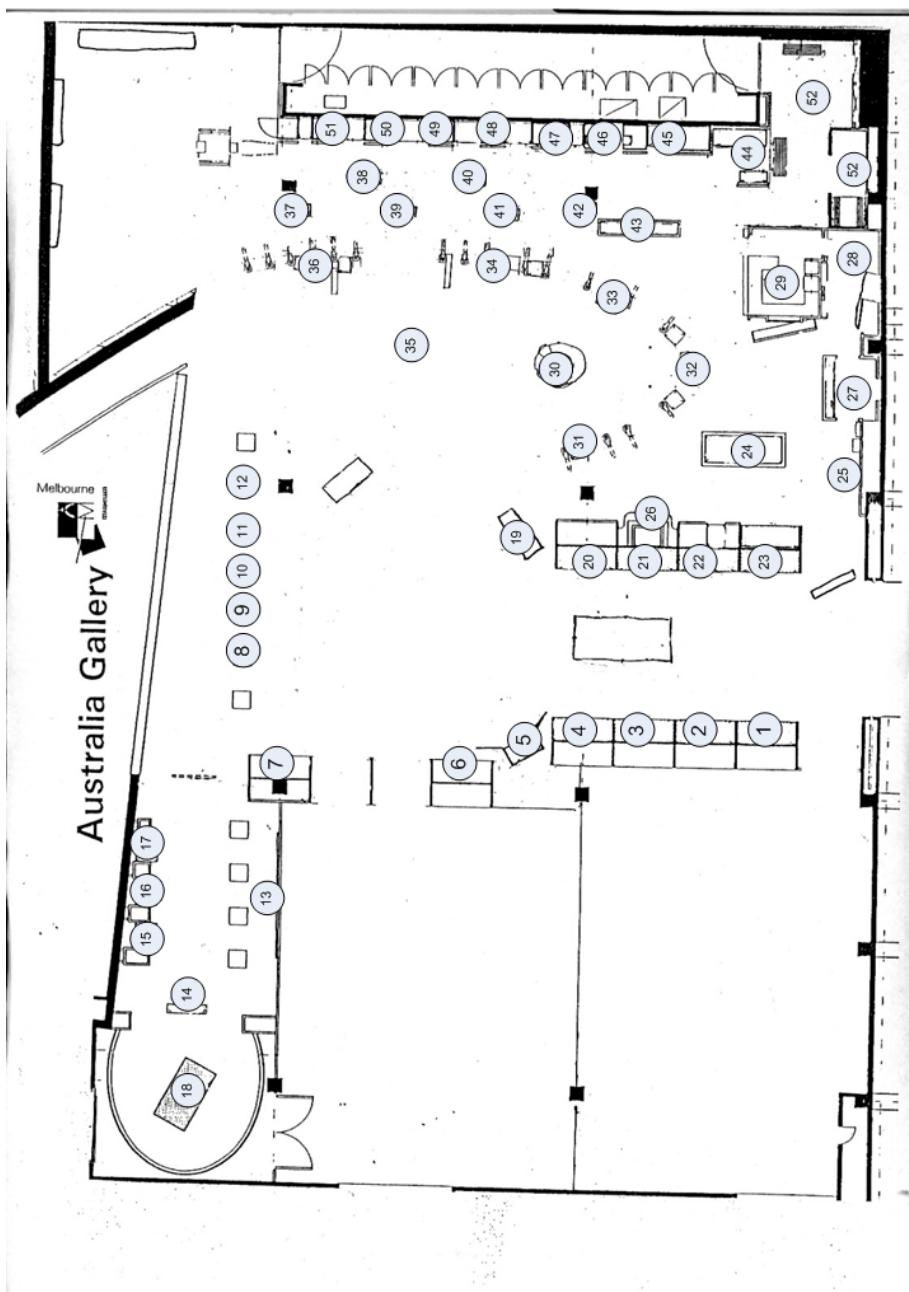


Figure 3.1: Map of the Australia Gallery, with exhibit positions labelled.

has not undergone large changes in the intervening five years. We assume that any changes made in the past five years have a minimal effect on the semantic ontology of the museum space and will not harm analysis of the collected visitor paths.

There are a total of sixty visitor paths recorded within the Australia Gallery. By combining the traced path of the visitors with the statistics recorded in the spreadsheet, we have created a corpus of sequential paths that can be thought of as sentences of exhibits with each exhibit representing a single word. These paths can be used to form a collaborative model of how visitors in general will behave within the environment. These paths will also be used as real world sample data to test the conceptual models developed.

### 3.2.3 Museum Website

The Melbourne Museum website<sup>3</sup> provides an overview of the museum, the research and collaborative projects associated with the museum, as well as additional information about the exhibitions within the museum. The content on the website is organised in the same fashion as the museum itself at the top level, with the content divided into the three major exhibition areas. At the lower level, the website groups together pages based on exhibition areas, and even goes as deep as having pages describing a single element of an exhibit display. This leads to areas of the web site that are highly similar, and closely align with the real world museum content that they are describing.

Along with the web pages describing the exhibits, there are available a number of information sheets that provide additional information on exhibits. These information sheets are available at the museum in hard copy, and provide information on topics covered by groups of exhibits. The information sheets are also available on the museum web site. They have been written by multiple authors, and hence they differ in style. A combination of web pages and information sheets were used to describe the content of each exhibit.

## 3.3 Natural Language Resources

A key component of this thesis is to demonstrate that natural language based conceptual models will outperform conceptual models that revolve around physical attributes of exhibits. As described in Chapter 4 one way of calculating similarity is by semantic similarity. The semantic similarity of two exhibits is how close the content of the exhibits are to one another. In order to process and analyse the information content associated with each exhibit, a range of Natural Language tools were used.

---

<sup>3</sup><http://www.museum.vic.gov.au>

### 3.3.1 WordNet

WordNet is an ontology for the English language used to represent the way in which lexical concepts are related. Within WordNet a concept is the lowest level itemset. A concept is an entity that describes a word, its meaning, and the relationships it shares with other concepts. The structure of the WordNet ontology is represented by lexical relationships such as synonyms, antonyms, hypernyms (generalisations) and hyponyms (specialisations). Its goal is to provide a combined thesaurus and dictionary that is usable for automatic text analysis. Words are classified into four distinct parts of speech: nouns, verbs, adjectives and adverbs. In this application WordNet is used to emulate visitor thought patterns, and to create a representation of how visitors form relationships between lexical content.

The common example used when describing WordNet is to look at the word *dog*. In this sense we are talking about the domesticated mammal. The meaning of *dog* in WordNet is given as:

a member of the genus *Canis* (probably descended from the common wolf)  
that has been domesticated by man since prehistoric times; occurs in many  
breeds; *the dog barked all night*

Hypernyms of *dog* include *canine* and *canid*, while hyponyms extend to specific breeds of dog. There also exist other relationships such as meronyms (part-of relationship: *dog* is a meronym of *pack*) and holonyms (made-of or has-part relationship: *dog* is a hyponym of *fur*). Using these relationships plus others present in WordNet allows us to identify similarities between words of which we may not have been previously aware.

WordNet (Fellbaum 1998) was created at Princeton University and is available for use by the public as well as limited commercial use. The version used in this project is WordNet v2.0.

### 3.3.2 WordNet::Similarity

Lexical relations describe how closely linked a given word pair is. There have been many measures created to gauge similarity or relatedness between words using the WordNet model (Jiang and Conrath 1997; Hirst and St-Onge 1998; Leacock *et al.* 1998; Lin 1998). These measures use differing lexical relationships between words to create a score based on a function of the distance between the two.

The WordNet::Similarity (Pedersen *et al.* 2004) package is a freely available set of Perl modules designed to provide a simple interface to WordNet. The interface determines the relatedness or similarity of two concepts across WordNet. WordNet::Similarity implements a total of nine measures using several of these measures. The three measures used in this project represent two chief methods of representing concept similarity (Lin 1998; Leacock *et al.* 1998) across WordNet, as well as

a relatedness measure based on the glossary of a word (Patwardhan and Pedersen 2003).

We used the WordNet::Similarity package over the key terms collected from the museum. The key terms describe the content associated with an exhibit, and so relationships between the key-terms equate to relationships between the exhibits themselves. Using the similarity measures available in WordNet::Similarity, the semantic relatedness between exhibits has been established and turned into a set of transitional probability matrices.

### 3.3.3 Lingpipe

Named Entity Detection and Recognition (NED/NER) is an active area of research in NLP. Named Entities are words or phrases that carry some meaning other than that implied by their dictionary definition. For example the Named Entity *John Smith* refers to a person by the name of John Smith, it is not just two words in a sentence. From this named entity we can infer in some way the subject of the body of text that it appears in.

Lingpipe is an application that deals with text processing and content identification and classification. Specifically it provides a means for identifying Named Entities within bodies of text.

The body of documents collected from the website (Section 3.2.3) has a large amount of information content. In order to extract meaningful information from the text surrounding it, we can use Named Entity Detection. Meaningful words and phrases can be extracted from the documents and phrases can be automatically extracted from the documents.

It is normally simple for a person to identify Named Entities in a body of text, due to prior knowledge of a phrases meaning or its placement within a sentence. However it is difficult to have a machine automatically extract Named Entities from text without a predefined list of such Entities. Unfortunately there is no list of meaningful terms available for the content of the museum website.

Automatic extraction of content using Lingpipe can be done using an Entity Detection method trained over news articles. The style of language used in this method is too far removed from the style of the museum web-pages. This results in entities only being partially recognised, or not at all in some cases. This necessitates that we create a method trained over documents of the same language style as those used in the web-pages. There is no ready made corpus of such documents, and so manual annotation of the web-pages would be needed to create an accurate classifier. Manual annotation negates the need for automatic classification.

In the end, this approach proved to be too arduous to be of any use. However it is worth noting that semantic relations created using Named Entities could conceivably prove more effective than the other method proposed in this research, and is worth future investigation, especially when dealing with larger bodies of documents.

# Chapter 4

## Methodology

The basis of this research is the creation of a recommender system in the form of a computational conceptual model designed to identify connections that visitors make between exhibits within a museum collection. This recommender system differs from traditional recommender systems in that it makes recommendations in a real world environment. Traditional recommender systems as in Resnick and Varian (1997) are based in an online environment, and do not take into account the effect that a physical environment has on user behaviour. In order to create a recommender system working within a physical space, environmental factors that can affect user behaviour need to be taken into account. The history of a visitor also plays a part in determining what related exhibits they might be interested in. In order to build a conceptual model for a visitor, knowledge of what a visitor has seen in the past is needed. Previous history provides a knowledge base of where the visitor currently is within the gallery, and what information is associated with the exhibits the visitor has already seen.

For a recommendation to be valid to a given visitor, it must be related to a visitor's conceptual model of the museum space. The information collected by the museum (Section 3.2.2) has been collated into a group of 60 paths, all represented by a sequence of exhibits that the visitor took an interest in. This information does not represent the conceptual model that the visitor adopts in the semantic space provided by the museum. In order to determine whether a recommended exhibit is an exhibit that the visitor does want to see, we analyse the total visitor history. After making a set of recommendations, we check each recommendation to see if it occurs within the visitor's history. If a recommended exhibit appears within this history then the recommendation is judged to be related to the conceptual model of the visitor. As previously stated, a recommendation can only be made if the visitor hasn't yet seen that exhibit. Recommendations are considered to be single instance events. If a recommendation is made to a visitor and they continue on their tour without visiting a recommendation, the visitor has either decided that the recommended exhibit was not relevant to their conceptual model, or decided that the recommendation was valid, but was distracted by another exhibit on the way to the recommended one.

Due to the nature of the visitor data collected, it is difficult to interpret which of these reasons is the contributing factor. The data is static, and will not respond to recommendations. For the purposes of the tests carried out within this research, results dealing with the prediction of exhibits must be taken with a grain of salt. The recommendations made will not alter the paths that the visitors take. The visitor will not have been directed to an exhibit, and may not be aware of its existence. However, if the visitor has encountered the exhibit further on in their path, it is valid to make the claim that they have found an exhibit relating to their predicted conceptual model. In this case we can claim that the initial recommendation was valid due to the fact that they have visited the exhibit after it has been recommended, and that it relates to their predicted conceptual model.

The induction of this recommendation validity model can be expressed as the precision of the recommendations. We wish to make correct recommendations, not a large number of recommendations that have a chance, however slight, of being relevant. This means that the quantity of recommendations made is not as important as the quality of the recommendations.

Each exhibit within the Australia Gallery has multiple levels of information associated with it. This is comprised of its physical location within the Gallery, the semantic information associated with it, and a related web page located on the Melbourne Museum website.

## 4.1 Physical Location

The Australia Gallery is a physical space and although we are using techniques adapted from hyperspace recommender systems (Resnick and Varian 1997), we need to recognise the impact the physical environment has on a visitor's ability to get from one exhibit to another. Visitors are less likely to travel directly to exhibits that are far away, and there may be other exhibits in the path of the visitor, blocking their view or travel route.

Physical proximity can be used as a measure of similarity between exhibits. It is a natural way of showing the importance the real world has on how visitors think about exhibit relationships. Exhibits closer to the current position of a visitor are more likely to be visited next, and visitors are unlikely to appreciate being recommended to travel to the opposite side of a collection. In a rigidly structured museum environment the physical closeness of an exhibit does not provide any information on what the thought processes of the visitor are. The visitors will simply progress from one exhibit to the next in a sequence as the visitor has little or no choice. In an open plan environment, the visitor is able to see the majority of the exhibits in the exhibition at once. The visitor is able to make a judgement about which exhibits most interest them from the selection they are able to see. The visitor is then more likely to choose an interesting exhibit that is closer than one on the other side of the room. Physical proximity is

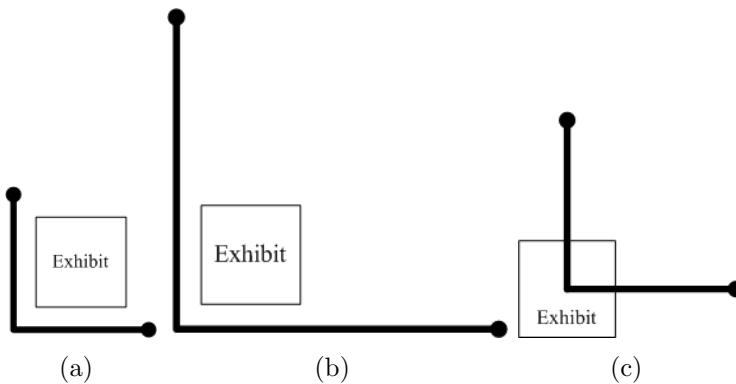


Figure 4.1: (a): Successful use of Manhattan Distance to avoid obstacles, (b): Taking the long way around, (c): Going through an obstacle

one contributing factor that can be used to model a visitor's thought process when making judgements.

A physical environment places constraints on visitor movement (Peponis *et al.* 2004), even in an open plan environment such as the Australia Gallery. Objects (seats, barriers, or other exhibits) prevent visitors from moving directly from one exhibit to another. To represent this limitation, the straight line distance from one exhibit to another cannot be used as an accurate measure of distance.

The measure of Manhattan Distance relates to movement in a two dimensional grid system. A path from one point on the grid to another can be found by moving in two straight lines: one movement East-West, and another North-South. The name of this measure comes from the Street-Avenue grid present on Manhattan Island in New York City. This measurement takes into account the obstacles that can be encountered when moving from one location to another, but is only accurate when the obstacles are arranged in a grid.

Unfortunately the exhibits within the Australia Gallery are not in a grid distribution, but are placed free-form around the exhibit space. Using Manhattan Distance in this format can lead to two problems: the shortest path taken goes through an intervening exhibit, or the path takes the long way around if there are no obstacles in the way.

In order to avoid going through obstacles, a graph representation of the environment can be created (Figure 4.2) to model the common intersections that paths have. This representation can be used by jumping from one node to another and finding the shortest path from the starting node to the end node. Within an environment where there are multiple paths from one location to another, the Shortest Path Problem becomes a factor. This can be solved using Dijkstra's algorithm to compute the shortest path from one node in a graph to another.

For the purposes of this experimentation, we use Manhattan Distance to describe

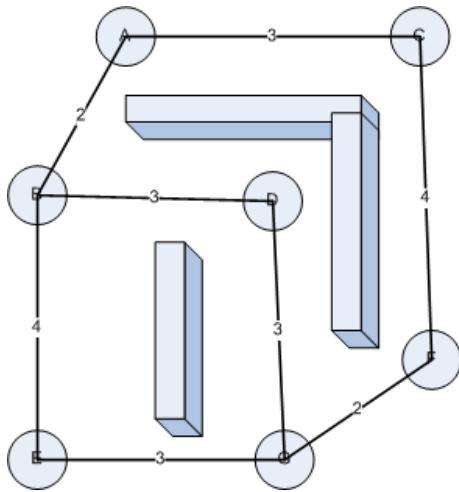


Figure 4.2: Graph representation of a physical environment with obstacles.

the distance from one exhibit to another. We acknowledge that this is a simplification of the problem, however the form of the physical locations used in the representation of each exhibit makes the use of the node-graph method infeasible. The use of other more complex distance metrics (node-graph representation) will be explored in future work (see Section 6.1).

The proximity similarity of any two exhibits is the Manhattan Distance between the two. This set of similarities is normalised into a transitional probability matrix

## 4.2 Collaborative Data

Many Recommender Systems use the opinions of previous users to give recommendations to new users (along the lines of Amazon.com’s “people who liked this book also liked...”). The sequential visitor data described in Section 3.2.2 can be used to form such a recommendation system. By expressing each visitor path as an ordered series of exhibits, we can determine how other future visitors are going to react in similar situations. If visitors do follow similar paths to previous visitors, either the visitors are following the structure of movement dictated by the exhibit layout or visitors are all attracted to similar items.

Collaborative methods make predictions based on the assumption that new users will behave in the same way as previous users, and the proviso of what is good for most will be good for all. This assumption works well if all users think alike, but users can rarely be lumped into a single overall characterisation. For the purposes of this research, we are not assuming one user type, or multiple user types, simply that there are users and that they may or may not follow the same movement patterns as previous visitors.

The methods of collaborative sequential data and physical proximity are superficial in scope and do not extend into the conceptual space adopted by the visitors. They do however give insight into how a physical space affects a visitors' mental representation of the conceptual areas associated with specific exhibit collections. Any accurate recommender systems produced in this fashion will need to take into account the limitations these two methods place on the thought processes of visitors.

### 4.2.1 Naive Bayes Prediction

When taking into account that a visitor's entire history will have an impact on what exhibit the visitor next visits, the quantity of training data is necessary to take into account. We can use an entire visitor path to predict the next exhibit, but for a given sequence, it must have occurred at least once before in the training data. If the sequence has not occurred, then the recommender will not be able to recommend an exhibit. We can avoid this problem of sparse data (only sixty visitor paths are available) by using the naive Bayes classifier. The naive Bayes classifier uses Bayes Theorem with the assumption that all events are independent, and have no effect on one another. Using the assumption that all conditional probabilities are independent to all other elements in the sequence, we can rewrite the classification as follows.<sup>1</sup>

$$P(c|A_1, \dots, A_n) = P(c) \prod_{i=1}^n P(A_i|c)$$

Now to select the most likely exhibit to go to next, we simply check the classification for all exhibits ( $c_i$ ) given the current history ( $A_1, \dots, A_t$ ):

$$\hat{c} = \arg \max_{c_i} P(c_i) \prod_{j=1}^t P(A_j|c_i) \times \frac{j}{t}$$

The final term is a scaling factor introduced to lessen the importance of exhibits that appeared earlier in the visitors path, where  $t$  is the total length of their path, and  $\hat{c}$  is the most likely next exhibit.

## 4.3 Semantic Content

The semantic content of an exhibit is described by a set of key-terms. These terms describe physical attributes of the exhibit, as well as the themes within an exhibit. The semantic similarity of two exhibits describes how close the exhibits are in terms of information content or physical appearance. An exhibit's semantic representation is designed to represent how a visitor thinks about an exhibit. When a visitor thinks of an exhibit that they have visited, they will recall the components that made up

---

<sup>1</sup>The full proof can be seen in Maron (1961)

the exhibit, or revisit an emotional reaction they may have had while looking at the exhibit. The similarity measures described later in this section are designed to infer the reason a visitor has for going from one exhibit to the next. For instance, if a visitor has spent a lot of time looking at exhibits to do with medical procedures, one may assume that they are interested in the human anatomy. However, if the visitor takes a keen interest in a display where carrion beetles are devouring an animal carcass,<sup>2</sup> the visitor may be interested in gory exhibits rather than medical exhibits. The problem lies in determining for what reason the visitor chose to visit a specific selection of exhibits.

### 4.3.1 Word Sense Disambiguation

We can describe a visitors path as a sentence of previously visited exhibits. With each exhibit in the path being a word in that sentence. The terms used to describe an exhibit can be considered different meanings (or senses) of the word. We can then use previously appearing exhibits in the visitor's path to identify why they visited a particular exhibit.

Word Sense Disambiguation (WSD) is an NLP problem occurring when extracting meaning from a body of text that contains words with multiple senses. The problem lies in determining which sense of the word is the correct one to infer. The context that the word appears in is key to disambiguating the word. Words that are related to one another are more likely to appear in the same context.

WSD methods can be adapted to identify the similarity between exhibits in a visitor path. Using the sentence-path analogy described above, we can consider the task of determining semantic similarity between exhibits as a WSD application. The definition of each exhibit is the set of terms that describe it. It is then a task of gauging similarity based on the definitions of each exhibit.

### 4.3.2 WordNet Similarity and Relatedness

WordNet provides a rich ontology of the English language, with each word having relationships and meanings. WordNet is a highly useful tool in WSD applications due to the fact that it represents relationships between words. If words appear together in a sentence, then the can supply information to help us disambiguate the meaning of other words in the sentence. There have been many methods designed to take advantage of WordNet's structure in WSD applications. The three we will use here provide a representative sample of different styles of measures. In WordNet 'similarity' refers to how close terms are with relation to the hierarchical structure of WordNet's IS-A (hypernym) and HAS-A (hyponym) relationships. These relationships are used to find the distance to a Least Common Subsumer (LCS), and compute the relative

---

<sup>2</sup>This exhibit actually exists within the Melbourne Museum.

path lengths from the terms to the LCS. Relatedness measures on the other hand rely on the definitions and glossaries associated with terms. By analysing similarities between the glossaries of words in a sentence, the most overlapping glossaries can indicate the correct word sense.

### **Leacock-Chodorow**

The Leacock-Chodorow method of similarity scoring uses the hierarchical structure of WordNet to find similarity between two words. The shortest path between two words is calculated using WordNet's hypernym and hyponym relationships. This shortest path is then scaled against the depth of the words within the hierarchy, i.e. their distance from the root node. The deeper a word concept is within the hierarchy, the more specific it is. *Corgi* is more specific than *animal* and is hence deeper in the hierarchy. If two terms low down in the hierarchy have a very short distance between them, then they will be highly similar.

### **Lin**

The Lin method of sense disambiguation (Lin 1998) also makes use of the hypernym-hyponym hierarchy of WordNet. The Lin method scores similarity based on commonality shared by word senses. The more commonality the two words share in their paths, the higher the similarity. The more differences they share, the lower their similarity. By describing the similarity of the paths to an LCS, the overall similarity of the two words can be effectively determined. The Lin method scales the Information Content of a word against its LCS with another word. With respect to its application within this research, the information content of exhibits will be scaled on how semantically diverse they are. If two exhibits contain a lot of detailed information, but aren't on the same topic, they will score low.

### **Banerjee-Pedersen**

The algorithm developed by Lesk calculates the relatedness of words using glossary similarities from a dictionary corpus. The variation on the Lesk algorithm described in Patwardhan and Pedersen (2003) uses WordNet relationships to compute the overlap between terms. Since a glossary defines a set of words with related meanings, the magnitude of the overlap between two glossaries can be used to determine how related two words are. Words that occur in definitions of a word sense as well as the same context as which the word is used are likely to be highly related. This relationship flows both ways, meaning that definitions of words help to disambiguate words in the same sentence or context. This expansion can be used to more accurately determine overlap between concepts. This method of sense disambiguation is intended to find similarities between the concepts represented within exhibits through information overlap.

## 4.4 Evaluating Component Effectiveness

The quality of the exhibits recommended can be interpreted as the number of correctly recommended exhibits versus the total number of recommendations. For the purposes of evaluation over the supplied visitor paths, an exhibit is correctly recommended if the exhibit appears in the visitor path after the point at which it was recommended. We perform this evaluation using the measures of Precision and Recall.

Precision and Recall are used as measures of success for applications of prediction or retrieval where the correct outcome is known beforehand. This allows one to interpret how successful a method was at its task as compared to a gold standard. These measures use the statistics of the True Positives and Negatives (TP and TN) and False Positives and Negatives (FP and FN) to calculate the correctness of a given classifier or prediction metric. The functions of precision and recall are:

$$\text{Precision} = \frac{TP}{TP + FP}, \text{Recall} = \frac{TP}{TP + FN}$$

These measures will return a value in the interval of zero to one. The precision of a given method is the number of correctly classified instances over the total number of instances classified. The recall of a measure is the number of correctly classified instances over the number of gold standard positive classifications. The measure of F-Score is a weighted harmonic average of the precision and recall:

$$F\text{-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

The use of Precision and Recall in the evaluation of Recommender Systems is an appropriate and valid application, and has been used effectively in studies such as Raskutti *et al.* (1997); Basu *et al.* (1998) (as cited in Zukerman and Albrecht (2001)) to evaluate predictive user modelling. The precision of a set of recommendations will give an indication of the quality of the recommendations returned. If the number of correct recommendations outweighs the number of incorrect recommendations, then the precision will be greater. In this case the recall of recommendations is less important. High recall indicates that a majority of the exhibits that a visitor visits are correctly recommended. This can also happen if many recommendations are made, meaning that the overall precision can drop if many incorrect recommendations are made. I.e. many correct recommendations are made, but also many incorrect ones. This will be detrimental to the success of the system. If many incorrect recommendations are made, the visitor will begin to ignore the recommendations after a while due to their overall irrelevance. If a high precision is achieved, this indicates that the system is able to correctly anticipate what a visitor is interested in, and hence forms a valid computational model of that user's thought patterns within the museum environment.

We describe the Accuracy of a recommender as the percentage of exhibits that the recommender correctly guessed would immediately succeed the currently visited exhibit. Accuracy is largely unused in this evaluation, as the sequence of exhibit visitation does not need to be predicted. The recommender functions adeptly if it is able to precisely predict exhibits that will be visited some time in the future.

# Chapter 5

## Experiments and Evaluation

Recommendation of similar exhibits can be modelled as a problem of predicting which exhibits are most likely to be seen. In this chapter predictive methods are created using the statistics of the gathered data detailed in Section 3.2. For a given visitor path, each predictive method produces a set of exhibits that are predicted to be relevant to the visitor’s path. Measures of physical proximity, semantic relatedness, and the paths of previous visitors are used to create these predictive methods. These methods are analysed individually, as well as in combination. Their effectiveness as methods of validly recommending exhibits, as well as their use in slightly more advanced predictive models is evaluated.

### 5.1 Experiment Design

Prediction of visitor paths can be expressed as selecting the most similar exhibit from a set of unvisited exhibits (as stated in Chapter 4). The experimentation in this chapter takes place over the collected visitor paths (Section 3.2.2). The predictive methods are tested by finding the exhibit most similar to a visitors prior path. For each exhibit a visitor has seen, a recommendation is made. For initial tests, the length of the recommendation set and the original path is identical. This results in identical Precision Recall and F-score for each method. These scores have been grouped under a single heading ‘Bag-of-exhibits’ (BOE). This describes the fact that these evaluations are made without considering the order of the recommendations. The measure of Accuracy is used to evaluate recommendations in the correct order, i.e. exhibits that were visited directly after they were recommended, not just some time in the future.

Method	BOE	Accuracy
Proximity	0.270	0.192
Collaborative	0.406	0.313
Cosine	0.187	0.070
Lin	0.129	0.039
Leacock-Chodorow	0.116	0.024
Banerjee-Pedersen	0.181	0.072

Table 5.1: Single Component Prediction Results

## 5.2 Single Component Predictive Methods

The predictions made in this section are made using only single similarity measures as discussed in Chapter 4. Evaluated are the effectiveness of physical similarity, previous visitor paths, cosine similarity, and the WordNet similarity measures described in Section 4.3.2.

From these results (Table 5.1) it can be seen that the similarity measure based on the paths of previous visitors is most effective in determining which exhibit a visitor will next see. Physical proximity manages to correctly identify the next exhibit to be visited approximately one out of every five times. This is not a highly successful method of prediction, but it indicates that visitors do not put as high an importance on physical similarity as previously assumed.

Recommendations made using cosine similarity across the semantic terms within the exhibits achieves a precision of 18.7%, closely followed by similarity calculated using the Banerjee-Pedersen Lesk algorithm. Both of these methods outdone by the (relative) success of the physical and collaborative methods. At first glance, this indicates that visitors do not make use of semantic based conceptual models when visiting a museum. However this conclusion does not take into account the application of this conceptual model to a physical environment. The constraints posed on the visitors by the environment contribute to the building of a valid conceptual model. Without the additional representation of the physical environment imposed on the conceptual model, recommendations from semantic based conceptual models will be incorrect more than 80% of the time.

### 5.2.1 Sequential Prediction

The naive Bayes method defined in Section 4.2.1 can be used in conjunction with any of the similarity measures above. The results of experimentation using the specified probability distributions are shown in Table 5.2.

These results are noticeably worse than the non-sequential predictive methods. This can mean that visitors do not make a full use of their history within the museum

NB-Method	BOE	Accuracy
Proximity	0.212	0.144
Collaborative	0.234	0.129
Cosine	0.151	0.055
Lin	0.144	0.057
Leacock-Chodorow	0.153	0.036
Banerjee-Pedersen	0.153	0.031

Table 5.2: Naive Bayesian Based Prediction Results

to make choices. Visitors use their immediate context to make their decision, rather than taking a considered approach to selecting their next exhibit. This indicates that visitors may not place specific importance on all items in their history, and only exhibits which the visitor found exceptional in some way will be remembered. With the collected visitor paths in their current state, it is very difficult to tell if a visitor has found an exhibit to be exceptional. The length of time that a visitor spends at an exhibit can provide a significant insight into how exceptional the visitor finds the exhibit. This is a factor that needs to be explored in future research.

### 5.3 Multiple Component Conceptual Models

The similarity between exhibits that form a visitor's conceptual model are made up of multiple components. The transitional matrices can be combined to produce predictive models that represent multiple aspects of the museum space. Combinations of transitional probabilities, as well as their effectiveness using the naive Bayes method is shown in Table 5.3.

The combination of multiple similarity measures are designed to express the multi-modal nature of the Australia Gallery. By combining probability distributions, we are able to express many qualities of an exhibit, and gain a more accurate picture of how visitors think about the relationships between them.

With regard to Tables 5.1 and 5.2, when a method is combined with the Collaborative method, its BOE score jumps significantly. The exception to this statement is Physical Proximity, this method when combined with Collaboration, comes out lower than as in Table 5.1. This indicates that people follow the same general paths as other people, but the exhibits in these paths are physically not that close to each other.

Semantic similarity measures in both the case of naive Bayes results and non-sequential results make a marginal improvement, but not significant enough to be noteworthy. The naive Bayes based methods continue to perform worse overall than their non-sequential counterparts.

Method Combination	BOE	Accuracy
Collaborative - Cosine	0.417	0.321
Collaborative - Proximity	0.223	0.130
Collaborative - Lin	0.361	0.271
Collaborative - Leacock-Chodorow	0.387	0.286
Collaborative - Banerjee-Pedersen	0.211	0.127
Proximity - Cosine	0.274	0.201
Proximity - Lin	0.237	0.154
Proximity - Leacock-Chodorow	0.250	0.154
Proximity - Banerjee-Pedersen	0.180	0.105
Naive Bayes Combinations		
Collaborative - Cosine	0.232	0.129
Collaborative - Proximity	0.226	0.157
Collaborative - Lin	0.225	0.114
Collaborative - Leacock-Chodorow	0.242	0.130
Collaborative - Banerjee-Pedersen	0.163	0.064
Proximity - Cosine	0.214	0.148
Proximity - Lin	0.180	0.114
Proximity - Leacock-Chodorow	0.220	0.151
Proximity - Banerjee-Pedersen	0.205	0.105

Table 5.3: Multiple Component Prediction Evaluation.

### 5.3.1 Improving Precision through Thresholds

The Precision Recall and F-score for all of the predictive methods are identical. This is due to the fact that the set of recommendations for a given visitor is the same length as their path. This results in the number of False negatives and False Positives being equal. To increase recommendation precision, the number of False Negatives must be reduced. The methods in section 5.2 will recommend an exhibit even if the transitional probability is very low. Low probability recommendations come from a visitor being at an exhibit that is highly unique, and does not share many similarities with other exhibits. This problem is more noticeable with semantic similarity as exhibits are more likely to be semantically different than for physical reasons. To prevent exhibits with low transitional probability being recommended, it is necessary to introduce a threshold that defines a minimum confidence that a transitional probability must achieve in order to be recommended.

A threshold will have the effect of reducing the recommendations recalled (a recommendation will not be made unless the certainty that it is relevant is high). Introducing a threshold will also increase the overall precision of a method. The threshold

Method	Threshold	Precision	Recall	F-Score
Single Component Prediction				
Proximity	0.03	0.271	0.270	0.270
Collaborative	0.04	0.471	0.283	0.354
Cosine	0.02	0.217	0.129	0.161
Lin	0.01	0.129	0.129	.0129
Leacock-Chodorow	0.01	0.117	0.117	0.117
Banerjee-Pedersen	0.01	0.182	0.180	0.181
Multiple Component Prediction				
Collaborative - Cosine	0.001	0.511	0.168	0.253
Collaborative - Proximity	0.001	0.262	0.144	0.186
Collaborative - Lin	0.0005	0.383	0.316	0.348
Collaborative - Leacock-Chodorow	0.0005	0.430	0.349	0.385
Collaborative - Banerjee-Pedersen	0.001	0.236	0.151	0.184
Proximity - Cosine	0.001	0.290	0.244	0.265
Proximity - Lin	0.0005	0.239	0.237	0.238
Proximity - Leacock-Chodorow	0.0005	0.252	0.250	0.251
Proximity - Banerjee-Pedersen	0.0005	0.182	0.180	0.181

Table 5.4: Results of predictive measures using thresholds to improve precision.

values were found using a manual hill-climbing approach to achieve maximum precision. Accuracy no longer plays a part as we are not recommending an ordered set or exhibits.

Using the methods described above thresholds have been introduced in order to increase their precision. The results of the threshold introduction are shown in Table 5.4. The small thresholds are because of the number of exhibits within the exhibition. All fifty-three exhibits within the Australia Gallery have a chance of being visited, and the average transitional probability is around 2% without any similarity taken into account. The multiple component conceptual models are even lower due to the fact that transitional probabilities are multiplied.

Using the requirement that we care about the quality of recommendations rather than the quantity, the introduction of thresholds greatly improves the results from our classifiers. The most precise method is the conjunction of prior visitor paths and cosine similarity of the semantic content associated with the exhibits. As in Section 5.3 all methods that use the prior visitor paths as a component have a markedly increased success rate. With thresholds introduced, the base precision of the individual Collaborative similarity measure is 47.1%. This in itself is a large improvement on the non-threshold results. When used in conjunction with other methods, this precision is reduced. Most notably when used in conjunction with the measure of

physical proximity, scoring lower than physical proximity alone.

We interpret this poor result as meaning that visitors generally follow the same paths as other visitors, but those paths are not affected by how close a given exhibit is. This indicates that visitors are not effected by the physical structure of the museum. From the combination of the Collaborative measure and the Cosine measure of semantic similarity, the highest Precision is achieved. Recommending a related exhibit more than 50% of the time. The best interpretation of this result is that visitors like what other visitors liked, but also place tend to remain on a similar theme to what they are currently viewing.

# Chapter 6

## Conclusion and Future Work

This research has shown that exhibit similarity based on the semantic content of the exhibits, and the exhibits contained in a visitor’s previous path can be used to form recommendations about which future exhibits the visitor may wish to see. As the museum is a real world environment, the visitor is effected by the physical constraints placed upon them by the structure of the exhibit. This can lead to multiple visitors following similar paths. Predictive methods based on the paths of previous visitors to the museum returned noticeably better recommendations than all other methods. Through the introduction of thresholds to produce a higher quality (if lower quantity) of recommendations resulted in the highest precision achieved. The combination of prediction based on the paths of previous visitors and augmented with cosine similarity produced a precision of 51.1%, when using a confidence threshold of 0.1%. Meaning that the method was able to give a successful recommendation more than 50% of the time.

More complex semantic similarity measures (Section 4.3.2) failed to score as highly as simple term matching (cosine similarity). This could be attributed to visitors not forming a complex conceptual model relating to each exhibit, and instead only examining superficial features of each exhibit. Physical proximity also failed to perform as expected, this can be explained by two elements: the open nature of the Australia Gallery in which the testing took place, and the nature of the distance measure used (see Section 4.1. Without a rigid exhibit layout to follow, visitor had the freedom to move about the collection however they pleased. The fact that the visitors still followed similar paths to previous visitors are attracted to the same exhibits for the same reason, whether they are eye catching, or just prominently placed within the collection.

## 6.1 Future work

The poor results when using physical proximity as the predictive method can be attributed to the distance metric used to calculate similarity. A distance metric that more correctly modelled the structure of the collection, would most likely have a marked improvement over the current results. The user themselves is another factor that was not taken into for this study. A reasonable hypothesis is that younger visitors are going to want to visit more interactive, or colourful exhibits, whereas older visitors are more likely to want information. The adaption of content presented to the visitors as in Hitzeman *et al.* (1997) is area of further expansion that could provide fascinating results. Visitors are more likely to respond positively to exhibits if the visitor is presented with information that they can easily enjoy, whether because it is aimed at their knowledge base or because the information has been adapted to include their own interests. Visitor feedback is most interesting area that has yet to be explored. If visitors can directly tell the recommender system what they want through approval or disapproval of the exhibits they have seen, highly specialised recommender systems can be made for individual visitors. Ideally these recommender systems will be tested out in an actual museum environment, providing recommendations and receiving feedback in real time.

## Acknowledgements

I wish to thank Tim Baldwin and Steven Bird for their supervision and guidance throughout this long (but funnily enough, too short) year; Carolyn Meehan and Alexa Reynolds for their help with gathering data and many ideas for projects; and to Ted Pedersen for developing that sterling package: WordNet::Similarity.

# Bibliography

BASU, CHUMKI, HAYM HIRSH, and WILLIAM COHEN. 1998. Recommendations as classification: Using social and content-based information in recommendation. In *Proceedings of the National Conference of Artificial Intelligence*, 714–720.

BENELLI, GIULIANO, ALBERTO BIANCHI, PATRIZIA MARTI, DAVID SENNATI, and ELENA NOT. 1999. HIPS: Hyper-Interaction within Physical Space. In *ICMCS '99: Proceedings of the IEEE International Conference on Multimedia Computing and Systems*, volume 2, p. 1075. IEEE Computer Society.

BITGOOD, S., and H. H. SHETTEL. 1997. An overview of visitor studies. *Journal of Museum Education* 21.6–10.

BRIN, SERGEY, and LAWRENCE PAGE. 1998. The anatomy of a large-sclae hyper-textual web search engine. In *WWW7: Proceedings of the seventh international conference on World Wide Web 7*, 107–117.

CHALMERS, MATTHEW. 2001. Paths and contextually specific recommendations. In *(DELOS) Workshop: Personalisation and Recommender Systems in Digital Libraries*.

—, KERRY RODDEN, and DOMINIQUE BRODBECK. 1998. The Order of Things: Activity-Centred Information Access. *Computer Networks and ISDN Systems* 30.1–7.

CHEVERST, K., and G. SMITH. 2001. Exploring the notion of information push and pull with respect to user intention and disruption. In *In International Workshop on Distributed and Disappearing User Interfaces in Ubiquitous Computing*, 67–72.

FELLBAUM, CHRISTIANE (ed.) 1998. *WordNet: An Electronic Lexical Database*. Cambridge, USA: MIT Press.

FILIPPINI, SILVIA. 2003. Personalisation through IT in museums: Does it really work?

HIRST, GRAEME, and DAVID ST-ONGE. 1998. Lexical chains as representations of context for the detection and correction of malapropism. In (Fellbaum 1998), 305–32.

HITZEMAN, JANET, CHRIS MELLISH, and JON OBERLANDER. 1997. Dynamic generation of museum web pages: The intelligent labelling explorer. *Archives and Museum Informatics* 11.

JIANG, JAY J., and DAVID W. CONRATH. 1997. Semantic similarity based on corpus statistics and lexical taxonomy. In *Proc. of ROCLING X*, Taipei, Taiwan.

LEACOCK, C., M. CHODROW, and G.A. MILLER. 1998. Using corpus statistics and WordNet relations for sense identification. *Computational Linguistics* 24.147–65.

LESK, MICHAEL. 1986. Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. In *SIGDOC '86: Proceedings of the 5th annual international conference on Systems documentation*, 24–26.

LIN, DEKANG. 1998. Automatic retrieval and clustering of similar words. In *(CoLING)-(ACL)*, 768–774.

MARON, M. E. 1961. Automatic indexing: An experimental inquiry. *J. ACM* 8.

PATWARDHAN, SIDDHARTH, and TED PEDERSEN. 2003. Extended gloss overlaps as a measure of semantic relatedness. In *International Joint Conference on Artificial Intelligence*, volume 18, 805–810.

PEDERSEN, TED, SIDDHARTH PATWARDHAN, and JASON MICHELIZZI. 2004. Wordnet::similarity - measring the relatedness of concepts. In *Demonstrations of the Human Language Technology Conference on Artificial Intelligence*, 1024–1025.

PEPONIS, JOHN, RUTH CONROY DALTON, JEAN WINEMAN, and NICK DALTON. 2004. Measuring the effect of layout on visitors' spatial behaviors in open plan exhibition settings. *Environment and Planning B: Planning and Design* 31.453–473.

RASKUTTI, BHAVANI, ANTHONY BEITZ, and BELINDA WARD. 1997. A feature-based approach to recommending selections based on past preferences. *User Modelling and Uaer Adaption* 7.179–218.

RESNICK, PAUL, and HAL R. VARIAN. 1997. Recommender systems. *Commun. ACM* 40.56–58.

SPARACINO, FLAVIA, KENT LARSON, RON MACNEIL, GLORIANNA DAVENPORT, and ALEX PENTLAND. 1999. Technologies and methods for interactive exhibit design: from wireless object and body tracking to wearable computers. In *International Cultural Heritage Informatics Meeting*.

ZUKERMAN, INGRID, and DAVID W. ALBRECHT. 2001. Predictive statistical models for user modeling. *User Modeling and User-Adapted Interaction* 11.5–18.

\$Id: thesis.tex,v 1.3 2006-11-03 04:13:05 tim Exp \$